第27卷 第3期
2010年06月

工 程 数 学 学 报
CHINESE JOURNAL OF ENGINEERING MATHEMATICS

Vol. 27 No. 3
June 2010

文章编号:1005-3085(2010)03-0429-12

含有变点的厚尾单位根的subsampling检验*

秦瑞兵[1]， 田 铮[1,2]

(1- 西北工业大学应用数学系，西安 710072;
2- 中国科学院遥感应用研究所国家遥感重点实验室，北京 100101)

摘 要: 本文研究趋势项含有变点且新息为方差无穷厚尾过程的序列单位根检验问题，通过构造DF型检验，得到了其渐近分布。为避免估计统计量渐近分布中的尾指数，构造subsampling抽样方法来确定统计量渐近分布的百分位数，同时论证了subsampling抽样方法的一致性。最后，用Monte Carlo模拟证实本文所提出统计量以及subsampling抽样方法的可行性。

关键词: 方差无穷过程；变点；单位根检验；subsampling方法

分类号: **AMS**(2000) 62F12 **中图分类号**: O212.1 **文献标识码**: A

1 引言

含有结构变点的单位根检验问题是近年来计量经济学与统计学界研究热点。Perron[1,2] 提出含有结构变点的单位根检验问题，并构造DF型检验统计量。Banerjee等人[3]，Christiano[4]，Zivot和Andrews[5] 考虑结构变点为内生变量时的单位根检验问题。Kim等[6] 考虑新息中存在方差变点时的单位根检验问题。Nunes[7] 用LM型检验研究了含有结构变点时的单位根检验问题。另一方面，Guillaume[8] 以及Mittnik和Rachev[9] 指出许多经济和金融数据具有尖峰厚尾的特点，用方差无穷序列来描述更能反应这些统计特性。Davis和Resnick[10,11] 给出方差无穷过程的样本相关函数以及样本协方差的收敛性。Chan和Tran[12] 针对新息为方差无穷厚尾过程情形下的单位根检验问题，使用DF方法(Dickey-Fuller)检验，结果表明直接使用DF检验，检验的势会有较大的波动。Phillips[13] 将这一结果推广到弱相依的方差无穷情形。在厚尾指数以及Chan和Tran[12] 中统计量渐近分布未知时，Agnieszka和Kokoszka[14] 应用subsampling方法确定上述DF统计量所需要的临界值，模拟结果表明subsampling方法有助于提高检验的势。

当新息为方差无穷过程时，对于在过程的趋势项含有结构变点时的单位根检验问题却鲜有研究。本文考虑了新息为方差无穷厚尾过程情形下，含有趋势变点时的单位根检验问题，得到此时DF检验的渐近分布。然而统计量的渐近分布结构较为复杂，且分布中含有尾指数，会对检验的势产生影响。为此，本文应用subsampling抽样方法来确定检验所需要的临界值，同时可以避免估计厚尾指数，并证明了subsampling抽样方法的一致性。数值模拟表明当样本容量较大时，其经验势函数接近1，这表明subsampling方法是有效的。

收稿日期: 2009-07-07. **作者简介**: 秦瑞兵(1979年8月生)，男，博士. 研究方向：金融时间序列.
***基金项目**: 国家自然科学基金(60972150; 10926197; 60375003)；西北工业大学科技创新基金(2007KJ01033).

2 模型与假设

考虑模型

$$y_t = \mu + \theta DU_t + \beta t + \rho y_{t-1} + u_t, \quad DU_t = \begin{cases} 1, & t \le k_0, \\ 0, & t > k_0, \end{cases} \quad t = 1, \cdots, n, \tag{1}$$

其中新息过程 $\{u_t\}$ 满足如下条件:

假设1 $\{u_t\}$ 为独立同分布的稳定分布序列，$Eu_t = 0$ 且其特征函数为

$$E\exp\{itu_1\} = \begin{cases} \exp\{-t^2/2\}, & \kappa = 2, \\ \exp\{-|t|^\kappa\}, & 1 < \kappa < 2. \end{cases}$$

检验原假设为 $H_0: \rho = 1$，备择假设为 $H_1: |\rho| < 1$。

对于上述单位根检验，由模型(1)得到 ρ 的最小二乘估计量 $\hat{\rho}_n$，从而构造DF检验 $T = n(\hat{\rho}_n - 1)$。由于最小二乘估计量 $\hat{\rho}_n$ 与模型参数 μ、θ、β 无关，因此在本文证明中假设:

假设2 $y_1, \cdots, y_n$ 的生成过程为 $y_t = y_{t-1} + u_t$，$t = 1, \cdots, n$，其中 u_t 满足假设1。

注1 为推导的简化，本节仅考虑新息为独立同分布的情形。事实上，本文结论对于更一般的弱相依厚尾序列也是成立的：$u_t = d(L)\varepsilon_t = \sum_{j=0}^{\infty} d_j \varepsilon_{t-j}$，其中系数 $d_0 = 1$，$d(1) \ne 0$，且满足 $\sum_{t=0}^{\infty} t|d_t|^\delta < \infty$，$0 < \delta < 1 \wedge \kappa$，误差 $\{\varepsilon_t\}$ 满足假设1。

3 主要结果

在假设1、2成立时，对DF检验 $T = n(1 - \hat{\rho}_n)$，有如下结果成立。

引理1 若 $\{y_t\}$ 为单位根过程，即 $y_t = y_t + u_t$，其中 u_t 满足假设1时，则有如下结论成立

$$a_n^{-1}\sum_{t=1}^{n} u_t \xrightarrow{d} U(1); \quad a_n^{-2}\sum_{t=1}^{n} y_t u_t \xrightarrow{d} \frac{1}{2}\left[U^2(1) - V(1)\right];$$

$$n^{-1}a_n^{-1}\sum_{t=1}^{n} t u_t \xrightarrow{d} U(1) - \int_0^1 U(r)\mathrm{d}r; \quad n^{-1}a_n^{-1}\sum_{t=1}^{n} y_{t-1} \xrightarrow{d} \int_0^1 U(r)\mathrm{d}r;$$

$$n^{-2}a_n^{-1}\sum_{t=1}^{n} t y_{t-1} \xrightarrow{d} \int_0^1 rU(r)\mathrm{d}r; \quad n^{-1}a_n^{-2}\sum_{t=1}^{n} y_{t-1}^2 \xrightarrow{d} \int_0^1 U^2(r)\mathrm{d}r;$$

$$n^{-1}a_n^{-1}\sum_{t=1}^{[n\lambda]} y_{t-1} \xrightarrow{d} \int_0^\lambda U(r)\mathrm{d}r; \quad n^{-1}a_n^{-2}\sum_{t=1}^{[n\lambda]} y_{t-1}^2 \xrightarrow{d} \int_0^\lambda U^2(r)\mathrm{d}r;$$

$$n^{-2}a_n^{-1}\sum_{t=1}^{[n\lambda]} t y_{t-1} \xrightarrow{d} \int_0^\lambda rU(r)\mathrm{d}r; \quad n^{-1}a_n^{-1}\sum_{t=1}^{[n\lambda]} t u_t \xrightarrow{d} \lambda U(\lambda) - \int_0^\lambda U(r)\mathrm{d}r.$$

证明 在假设1下，对 $0 < r < 1$，有 $a_n^{-1}\sum_{t=1}^{[nr]} u_t \xrightarrow{d} U(r)$，结合文献[15]的证明，可以证明引理1的结论成立。

定理1 在假设1中，当原假设 H_0 成立且变点时刻 $k_0 = [n\lambda]$ 已知时，$\rho = 1$ 的最小二乘估计 $\hat{\rho}_n$ 满足

$$n(\hat{\rho}_n - 1) \xrightarrow{d} \frac{(dc - b^2)D_1 - cD_2 + bD_3 + dD_4 + bD_5}{(dc - b^2)D_6 - dD_7 - cD_8 + 2bD_9}, \tag{2}$$

其中 $d=1/12$，$b=-\frac{(1-\lambda)\lambda}{2}$，$c=(1-\lambda)\lambda$.

$$D_1=\frac{1}{2}\left[U^2(1)-V(1)\right]-U(1)\int_0^1 U(r)\mathrm{d}r,\quad D_5=\big(U(\lambda)-\lambda U(1)\big)\int_0^1\Big(r-\frac{1}{2}\Big)U(r)\mathrm{d}r,$$

$$D_2=\Big(\frac{1}{2}U(1)-\int_0^1 U(r)\mathrm{d}r\Big)\int_0^1\Big(r-\frac{1}{2}U(r)\Big)\mathrm{d}r,\quad D_6=\int_0^1 U^2(r)\mathrm{d}r-\Big(\int_0^1 U(r)\mathrm{d}r\Big)^2,$$

$$D_3=\Big(\int_0^\lambda U(r)\mathrm{d}r-\lambda\int_0^1 U(r)\mathrm{d}r\Big)\Big(\frac{1}{2}U(1)-\int_0^1 U(r)\mathrm{d}r\Big),\quad D_7=\Big(\int_0^\lambda U(r)\mathrm{d}r-\lambda\int_0^1 U(r)\mathrm{d}r\Big)^2,$$

$$D_4=\Big(\int_0^\lambda U(r)\mathrm{d}r-\lambda\int_0^1 U(r)\mathrm{d}r\Big)\big(U(\lambda)-\lambda U(1)\big),\quad D_8=\Big(\int_0^1\Big(r-\frac{1}{2}\Big)U(r)\mathrm{d}r\Big)^2,$$

$$D_9=\Big(\int_0^\lambda U(r)\mathrm{d}r-\lambda\int_0^1 U(r)\mathrm{d}r\Big)\int_0^1\Big(r-\frac{1}{2}\Big)U(r)\mathrm{d}r,$$

其中 $U(r)$ 为具有特征指数 κ 的稳定分布变量。

证明　为证明的方便，将对假设1的回归分为两步：第一步，对模型各项的常数作回归得到

$$Y_t=\beta X_{1t}+\theta X_{2t}+\rho\hat{y}_{t-1}+u_t, \tag{3}$$

其中

$$Y_t=y_t-\frac{1}{n}\sum_{t=1}^n y_t,\quad \hat{y}_{t-1}=y_{t-1}-\frac{1}{n}\sum_{t=1}^n y_{t-1},$$

$$X_{1t}=t-\frac{1}{n}\sum_{t=1}^n t=t-\frac{n+1}{2},\quad X_{2t}=DU_t-\lambda,\quad Y_{-1}=\frac{1}{n}\sum_{t=1}^n y_{t-1}.$$

若记

$$Y'=(\hat{y}_0,\cdots,\hat{y}_{t-1}),\quad U'=(u_1,\cdots,u_n),\quad Z'=\left(\begin{array}{c}X_{11},\cdots,X_{1n},\\ X_{21},\cdots,X_{2n},\end{array}\right).$$

第二步，对模型(3)进行回归，得到 $\hat{\rho}_n$，并作如下分解

$$\hat{\rho}_n-1=(Y'Y)^{-1}Y'E-(Y'Y)^{-1}Z\big[Z'(I-P_Y)Z\big]^{-1}Z'(I-P_Y)E, \tag{4}$$

其中 $P_Y=Y(Y'Y)^{-1}Y'$。若令

$$Z'Z=\begin{pmatrix}d_n & b_n\\ b_n & c_n\end{pmatrix},\quad Z'Y=\begin{pmatrix}H\\ J\end{pmatrix},\quad Z'E=\begin{pmatrix}K\\ L\end{pmatrix},$$

则(4)式可以表示为

$$\hat{\rho}_n-1=\frac{(d_nc_n-b_n^2)Y'E-c_nHK+b_nJK-d_nJL+b_nHL}{(d_nc_n-b_n^2)Y'Y-d_nJ^2-d_nJ^2-c_nH^2+2b_nHJ}. \tag{5}$$

结合 d_n, c_n, b_n, H, J, K, L 定义，由引理1可以得到(4)式右端各项的极限如下

$$n^{-3}d_n = n^{-3}(n^3/12 - n/12) \to d, \quad n^{-2}b_n = -n^{-2}(1-\lambda)\lambda n^2/2 \to b,$$

$$n^{-1}c_n = n^{-1}(1-\lambda)\lambda n \to c, \quad a_n^{-1}L = \sum_{t=1}^{T_B}(u_{t-1} - \bar{u}) \xrightarrow{d} U(\lambda) - \lambda U(1).$$

$$a_n^{-2}Y'E = a_n^{-2}\sum_{t=1}^{n}(y_{t-1} - Y_{-1})u_t \xrightarrow{d} \frac{1}{2}\left[U^2(1) - V(1)\right] - U(1)\int_0^1 U(r)\mathrm{d}r,$$

$$n^{-1}a_n^{-2}Y'Y = n^{-1}a_n^{-2}\sum_{t=1}^{n}(y_{t-1} - Y_{-1})^2 \xrightarrow{d} \int_0^1 U^2(r)\mathrm{d}r - \left(\int_0^1 U(r)\mathrm{d}r\right)^2,$$

$$n^{-2}a_n^{-1}H = n^{-2}a_n^{-1}\sum_{t=1}^{n}t(y_t - y_{-1}) \xrightarrow{d} \int_0^1 rU(r)\mathrm{d}r - \frac{1}{2}\int_0^1 U(r)\mathrm{d}r,$$

$$n^{-1}a_n^{-1}K = n^{-1}a_n^{-1}\sum_{t=1}^{n}\left(t - \frac{n+1}{2}\right)u_t \xrightarrow{d} \frac{1}{2}U(1) - \int_0^1 U(r)\mathrm{d}r,$$

$$n^{-1}a_n^{-1}J = n^{-1}a_n^{-1}\sum_{t=1}^{T_B}(y_{t-1} - Y_{-1}) \xrightarrow{d} \int_0^\lambda U(r)\mathrm{d}r - \lambda\int_0^1 U(r)\mathrm{d}r.$$

对DF型统计量 $n(\hat{\rho}_n - 1)$，我们有

$$\begin{aligned} n(\hat{\rho}_n - 1) &= \frac{(d_nc_n - b_n^2)Y'E - c_nHK + b_nJK - d_nJL + b_nHL}{(d_nc_n - b_n^2)Y'Y - d_nJ^2 - d_nJ^2 - c_nH^2 + 2b_nHJ} \\ &\xrightarrow{d} \frac{(dc - b^2)D_1 - cD_2 + bD_3 + dD_4 + bD_5}{(dc - b^2)D_6 - dD_7 - cD_8 + 2bD_9}, \end{aligned}$$

定理证毕。

由定理1可知，当新息过程 u_t 为方差无穷过程时，DF型检验统计量 $n(\hat{\rho}_n - 1)$ 的渐近分布为Lévy过程的泛函，因此临界值的选取会依赖于对新息过程中的厚尾指数 κ 的精确估计。为避免估计 κ，我们用subsampling方法来逼近DF检验的渐近分布，从而将抽样的经验分布分位数作为DF型检验的临界值，具体步骤为：

1) 计算残差 $\hat{u}_k = y_k - \hat{\rho}_n y_{k-1}$，$k = 2, \cdots, n$，并将其中心化，且记为

$$u_k^0 = \hat{u}_k - \frac{1}{n-1}\sum_{t=2}^{n}\hat{u}_t;$$

2) 基于上述中心化后的残差过程 u_k^0，构造 $n-b$ 个子过程，当 $k = 2, \cdots, n-b+1$，第 k 个子过程定义为 $y_0(k) = 0$，$y_1(k) = u_k^0$，$y_2(k) = y_1(k) = u_{k+1}^0$，$\cdots$，$y_b = y_{b-1} + u_{k+b-1}^0$；

3) 记 $\hat{\rho}_{b,k}^0$ 为基于子过程 $y_0(k), \cdots, y_b(k)$ 在模型(1)中回归所得关于 ρ 的估计，记 $T_{b,k}^0, k = 2, \cdots, n-b+1$，所生成经验分布的 α 分位数为 $q_{n,b}^0(\alpha)$，若 $n(\hat{\rho}_n - 1) < q_{n,b}^0(\alpha)$，则拒绝原假设。

记统计量 $n(\hat{\rho}_n - 1)$ 的渐近分布函数为 $G(x)$，则对上述subsampling方法得到的经验分布函数 $\hat{G} = \frac{1}{n-b}\sum_{k=2}^{n-b+1} I\{T_{b,k}^0 \le x\}$ 有如下的定理成立。

定理2 在原假设下，当 $n \to \infty$，$b \to \infty$ 且 $b/n \to 0$ 时，对任意 x，均有 $\hat{G}(x) \xrightarrow{P} G(x)$。

定理3　在备则假设下，当$n\to\infty$, $b\to\infty$且$b/n\to 0$时，对任意x，均有

$$\hat{G}(x)\xrightarrow{P} H(x)\equiv 1.$$

注2　定理2表明在原假设成立时，用subsampling方法确定的临界值趋于真实分布的临界值G。定理3则表明在备则假设成立时，用subsampling抽样方法确定的临界值进行检验，其经验势函数接近于1。

为方便定理2的证明，先证明如下的引理2至引理5。

引理2　设随机变量序列S_n，$U_{k,n}$，$1\le k\le n$，满足$S_n\to 0$, $U_{k,n}\overset{d}{=}U_{1,n}$且$U_{1,n}=O_P(1)$，则对任意$\epsilon>0$, $\limsup\limits_{n\to\infty}\max\limits_k P(|S_nU_{k,n}|>\epsilon)=0$。

引理2的证明参见文献[14]。

引理3　在定理2的条件下，对于$k=2,\cdots,n$, $u_k^0=u_k+r_k(n)$，且$r_k(n)=O_P(n^{-1}a_n)$。

证明　由$r_k(n)$的定义可知

$$\begin{aligned}
r_k(n) &= u_k^0-u_k=\hat{u}_k-\frac{1}{n-1}\sum_{t=2}^{n}\hat{u}_t-u_k\\
&= [y_k-\hat{\rho}_n y_{k-1}]-\frac{1}{n-1}\sum_{t=2}^{n}[y_t-\hat{\rho}_n y_{t-1}]-u_k\\
&= \big[y_k-y_{k-1}-(\hat{\rho}_n-1)y_{k-1}\big]-\frac{1}{n-1}\sum_{t=2}^{n}\big[y_t-y_{t-1}-(\hat{\rho}_n-1)y_{t-1}\big]-u_k\\
&= -\frac{1}{n-1}\sum_{t=2}^{n}u_k-\frac{1}{n}n(\hat{\rho}_n-1)\Big(y_{k-1}-\frac{1}{n-1}\sum_{t=2}^{n}y_{t-1}\Big),
\end{aligned}$$

由于

$$\frac{1}{n-1}\sum_{t=2}^{n}u_t=O_P(n^{-1}a_n),\quad \frac{1}{n-1}y_{k-1}=O_P(n^{-1}a_n),$$

$$\frac{1}{n(n-1)}\sum_{t=2}^{n}y_{k-1}=O_P(n^{-1}a_n),\quad n(\hat{\rho}-1)=O_P(1),$$

从而$r_k=O_P(n^{-1}a_n)$，引理证毕。

由于在后续引理的证明中，将作类似定理1证明中的分解，故为方便，记$a_{b,k}=b^3/12-b/12$, $b_{b,k}=-(1-\lambda)\lambda b^2/2$, $c_{b,k}=(1-\lambda)\lambda b$.

$$Y'E_{b,k}=\sum_{i=k}^{k+b-1}y_{i-1}u_i-\sum_{i=k}^{k+b-1}u_i\Big(\frac{1}{b}\sum_{j=k}^{k+b-1}y_j\Big),\quad H_{b,k}=\sum_{i=k}^{k+b-1}i\Big[y_{i-1}-\frac{1}{b}\sum_{j=k}^{k+b-1}y_{j-1}\Big],$$

$$K_{b,k}=\sum_{i=k}^{k+b-1}\Big(iu_i-\frac{b+1}{2}\sum_{j=k}^{k+b-1}u_j\Big),\quad J_{b,k}=\sum_{i=k}^{k+\lambda b-1}\Big[y_{i-1}-\Big(\frac{1}{b}\sum_{j=k}^{k+b-1}y_{j-1}\Big)\Big],$$

$$L_{b,k}=\sum_{i=k}^{k+\lambda b-1}\Big(u_i-\frac{1}{b}\sum_{j=k}^{k+b-1}u_j\Big),\quad Y'Y_{b,k}=\sum_{i=k}^{k+b-1}\Big[y_{i-1}-\frac{1}{b}\sum_{j=k}^{k+b-1}y_{j-1}\Big]^2,$$

$$Y'E^0_{b,k}=\sum_{i=k}^{k+b-1}y_{i-1}(k)u_i^0-\sum_{i=k}^{k+b-1}u_i^0\Big[\frac{1}{b}\sum_{j=k}^{k+b-1}y_{j-1}(k)\Big],$$

$$H^0_{b,k}=\sum_{i=k}^{k+b-1}i\Big[y_{i-1}(k)-\Big(\frac{1}{b}\sum_{j=k}^{k+b-1}y_{j-1}(k)\Big)\Big],\quad K^0_{b,k}=\sum_{i=k}^{k+b-1}\Big(iu_i^0-\frac{b+1}{2}\sum_{i=k}^{k+b-1}u_i^0\Big),$$

$$J^0_{b,k}=\sum_{i=k}^{k+\lambda b-1}i\Big[y_{i-1}(k)-\frac{1}{b}\sum_{j=k}^{k+b-1}y_{j-1}(k)\Big],\quad L^0_{b,k}=\sum_{i=k}^{k+\lambda b-1}\Big(u_i^0-\frac{1}{b}\sum_{i=k}^{k+b-1}u_i^0\Big),$$

$$Y'Y^0_{b,k}=\sum_{i=k}^{k+b-1}\Big[y_{i-1}(k)-\frac{1}{b}\sum_{j=k}^{k+b-1}y_{i-1}(k)\Big]^2,$$

$$\begin{aligned}T^N_{b,k}=b^{-4}a_b^{-2}\big[&(a_{b,k}c_{b,k}-b^2_{b,k})Y'E_{\mathrm{b,k}}-c_{b,k}H_{\mathrm{b,k}}K_{\mathrm{b,k}}+b_{b,k}J_{\mathrm{b,k}}K_{\mathrm{b,k}}\\&-a_{b,k}J_{\mathrm{b,k}}L_{\mathrm{b,k}}+b_{b,k}H_{b,k}L_{b,k}\big],\end{aligned}$$

$$T^D_{b,k}=b^{-5}a_b^{-2}[(a_{b,k}c_{b,k}-b^2_{b,k})Y'Y_{\mathrm{b,k}}-a_{b,k}J_{\mathrm{b,k}}J_{\mathrm{b,k}}-c_{b,k}H_{\mathrm{b,k}}H_{\mathrm{b,k}}+2b_{b,k}H_{\mathrm{b,k}}J_{\mathrm{b,k}}],$$

类似可定义$T^{0,N}_{b,k}$与$T^{0,D}_{b,k}$，从而$n(\hat\rho_n-1)=T^N_{b,k}/T^D_{b,k}$，$b(\hat\rho^0_{b,k}-1)=T^{0,N}_{b,k}/T^{0,D}_{b,k}$，$k=2,\cdots,n-b+1$。在这些条件下，可以证明。

引理4 在定理2的条件下，对任意x，均有$\lim\limits_{n\to\infty}\frac{1}{n-b}\sum_{k=2}^{n-b+1}P(T^0_{b,k}\le x)\to G(x)$。

证明 在证明前，为方便，定义两个量：$R^N_{b,k}=T^{0,N}_{b,k}-T^N_{b,k}$，$R^D_{b,k}=T^{0,D}_{b,k}-T^D_{b,k}$。注意到事件$\{T^0_{b,k}\leqslant x\}$与事件$\{T^N_{b,k}+R^N_{b,k}\leqslant xT^D_{b,k}+xN^D_{b,k}\}$等价，从而可得

$$P(T^0_{b,k}\leqslant x)\leqslant P(T^N_{b,k}\leqslant xT^D_{b,k}+2\varepsilon)+P(R^N_{b,k}\leqslant-\varepsilon)+P(xR^D_{b,k}\geqslant\varepsilon),\tag{6}$$

$$P(T^0_{b,k}\leqslant x)\geqslant P(T^N_{b,k}\leqslant xT^D_{b,k}-2\varepsilon)-P(R^N_{b,k}>\varepsilon)+P(xR^D_{b,k}\leqslant-\varepsilon).\tag{7}$$

若能证明

$$\limsup_{n\to\infty}\max_{2\leqslant k\leqslant n-b+1}P\big\{|R^N_{b,k}|>\varepsilon\big\}=0,\quad\limsup_{n\to\infty}\max_{2\leqslant k\leqslant n-b+1}P\big\{|R^D_{b,k}|>\varepsilon\big\}=0,\quad\forall\epsilon>0,$$

则由上、下极限的性质以及引理3可得

$$\lim_{n\to\infty}\frac{1}{n-b}\sum_{k=2}^{n-b+1}P(T^0_{b,k}\leqslant x)\to G(x),$$

从而引理证毕。下仅证

$$\limsup_{n\to\infty}\max_{2\leqslant k\leqslant n-b+1}P\big\{|R^N_{b,k}|>\varepsilon\big\}=0,$$

对$R^D_{b,k}$可类似处理。记

$$S_{\mathrm{b,k}}(t)=a_{\mathrm{b}}^{-1}\sum_{i=k}^{[(k+b-1)t]}r_i(n),$$

则由引理3，结合$b/n\to0$，$1<\kappa<2$，有

$$\sup_{0\leqslant t\leqslant1}\max_{1\leqslant k\leqslant n-b+1}|S_{b,k}(t)|=O_P\Big(\frac{b}{a_b}\cdot\frac{a_T}{T}\Big)=o_P(1).\tag{8}$$

记 $R_{b,k}^N = T_{b,k}^{0N} - T_{b,k}^N = U - V + X - Y + Z$，其中 U, V, X, Y, Z 定义如下

$$V = b^{-4}a_b^{-2}c_{b,k}H_{b,k}^0K_{b,k}^0 - b^{-4}a_b^{-2}c_{b,k}H_{b,k}K_{b,k},$$

$$X = b^{-4}a_b^{-2}b_{b,k}J_{b,k}^0L_{b,k}^0 - b^{-4}a_b^{-2}b_{b,k}J_{b,k}L_{b,k},$$

$$Y = b^{-4}a_b^{-2}a_{b,k}J_{b,k}^0L_{b,k}^0 - b^{-4}a_b^{-2}a_{b,k}J_{b,k}L_{b,k},$$

$$Z = b^{-4}a_b^{-2}a_{b,k}H_{b,k}^0L_{b,k}^0 - b^{-4}a_b^{-2}a_{b,k}H_{b,k}L_{b,k},$$

$$\begin{aligned}U &= b^{-4}a_b^{-2}(a_{b,k}c_{b,k} - b_{b,k}^2)Y'E_{b,k}^0 - b^{-4}a_b^{-2}(a_{b,k}c_{b,k} - b_{b,k}^2)Y'E_{b,k}\\ &= b^{-4}a_b^{-2}[Y'E_{b,k}^0 - Y'E_{b,k}],\end{aligned}$$

由上述 $Y'E_{b,k}^0$ 与 $Y'E_{b,k}$ 的定义可得

$$\begin{aligned}&a_b^{-2}\big[Y'E_{b,k}^0 - Y'E_{b,k}\big]\\ &= a_b^{-2}\Big(\sum_{i=k}^{k+b-1} y_{i-1}(k)u_i^0 - \sum_{i=k}^{k+b-1} y_{i-1}(k)u_i\Big)\\ &= \Big(a_b^{-1}\sum_{i=k}^{k+b-1} u_i\Big)S_{b,k}(1) + \frac{1}{2}S_{b,k}^2(1) - a_b^{-2}\sum_{i=k}^{k+b-1} u_ir_i(n) - \frac{1}{2}a_b^{-2}\sum_{i=k}^{k+b-1} r_i^2(n),\end{aligned}$$

从而由(8)式可得

$$\Big(a_b^{-1}\sum_{i=k}^{k+b-1} u_i\Big)S_{b,k}(1) = o_P(1),$$

结合引理3可得，$a_b^{-2}[Y'E_{b,k}^0 - Y'E_{b,k}] = o_P(1)$，因此 $U = o_P(1)$。类似可以证明 $V = o_P(1)$，$X = o_P(1)$，$Y = o_P(1)$，$Z = o_P(1)$，综合则有

$$\limsup_{n\to\infty}\max_{2\leqslant k\leqslant n-b+1} P\big\{|R_{b,k}^N| > \varepsilon\big\} = 0.$$

引理5 在定理2的条件成立时，对任意 x，均有 $\mathrm{Var}[\hat{G}_b(x)] \to 0$。

证明 由 $\hat{G}_b(x)$ 的定义有

$$\begin{aligned}\mathrm{Var}\big[\hat{G}_b(x)\big] &= \frac{1}{(n-b)^2}\sum_{k=1}^{n-b+1}\mathrm{Var}\big[I\{T_{b,k}^0 \leqslant x\}\big]\\ &\quad+\frac{2}{(n-b)^2}\sum_{j=1}^{n-b}\sum_{k=1}^{n-b-j}\mathrm{Cov}\big(I\{T_{b,k}^0 \leqslant x\}, I\{T_{b,k+j}^0 \leqslant x\}\big)\\ &= S_1 + S_2.\end{aligned}$$

由引理4的证明可知，$S_1 \to 0$。将 S_2 作如下分解

$$\begin{aligned}S_2 &= \frac{2}{(n-b)^2}\sum_{j=1}^{b}\sum_{k=1}^{n-b-j}\mathrm{Cov}\big(I\{T_{b,k}^0 \leqslant x\}, I\{T_{b,k+j}^0 \leqslant x\}\big)\\ &\quad+\frac{2}{(n-b)^2}\sum_{j=b+1}^{n-b}\sum_{k=1}^{n-b-j}\mathrm{Cov}\big(I\{T_{b,k}^0 \leqslant x\}, I\{T_{b,k+j}^0 \leqslant x\}\big)\\ &= S_{21} + S_{22}.\end{aligned}$$

由 $b/n \to 0$ 可得 $S_{21} \to 0$，下证 $S_{22} \to 0$。由引理4的证明过程可知

$$\delta_n = \max_{1 \leqslant k \leqslant n-b+1} \left[P\{|R_{b,k}^N| \geqslant \varepsilon\} + P\{|xR_{b,k}^D| \geqslant \varepsilon\} \right] \to 0.$$

定义 $p_n^+ = P(T_{b,k}^N \leqslant xT_{b,k}^D + 2\varepsilon)$, $p_n^- = P(T_{b,k}^N \leqslant xT_{b,k}^D - 2\varepsilon)$,

$$p_{n,j}^+ = P\big(T_{b,k}^N \leqslant xT_{b,k}^D + 2\varepsilon,\ T_{b,k+j}^N \leqslant xT_{b,k+j}^D + 2\varepsilon \big),$$

$$p_{n,j}^- = P\big(T_{b,k}^N \leqslant xT_{b,k}^D + 2\varepsilon,\ T_{b,k+j}^N \leqslant xT_{b,k+j}^D - 2\varepsilon \big).$$

由(6)与(7)式可得

$$p_n^- - \delta_n \leqslant P(T_{b,k}^0 \leqslant x) \leqslant p_n^+ + \delta_n,$$

$$p_{n,j}^- - 2\delta_n \leqslant P\big(T_{b,k}^0 \leqslant x, T_{b,k+j}^0 \leqslant x \big) \leqslant p_{n,j}^+ + 2\delta_n,$$

因此

$$\mathrm{Cov}\big(I\{T_{b,k}^0 \leqslant x\}, I\{T_{b,k+j}^0 \leqslant x\} \big) \leqslant p_{n,j}^+ - (p_n^-)^2 + 2\delta_n + \delta_n^2,$$

从而

$$S_{22} \leqslant \frac{2}{(n-b)^2} \sum_{j=b+1}^{n-b} (n-b-j-1)\big\{ p_{n,j}^+ - (p_n^-)^2 + 2\delta_n + \delta_n^2 \big\},$$

故

$$\limsup_{n\to\infty} S_{22} \leqslant 2\big[(p^+)^2 - (p^-)^2 \big], \quad p^+ = \lim_{n\to\infty} p_n^+, \quad p^- = \lim_{n\to\infty} p_n^-.$$

由 $\epsilon > 0$ 的任意性可得 $\limsup\limits_{n\to\infty} S_{22} = 0$。

定理2的证明 为证 $\hat{G}_b(x) \xrightarrow{P} G(x)$，仅需证明 $E[\hat{G}_b(x) - G(x)] \to 0$, $\mathrm{Var}[\hat{G}_b(x) - G(x)] \to 0$。而由引理4、5可以得到 $E[\hat{G}_b(x) - G(x)] \to 0$, $\mathrm{Var}[\hat{G}_b(x) - G(x)] \to 0$，从而定理证毕。

为证明定理3，先证明如下3个引理。

引理6 在定理3的条件下，有 $u_k^0 = u_k + \gamma_k + \eta(n)$，此处

$$\gamma_k = -(\hat{\rho}_n - \rho) y_{k-1}, \quad \eta(n) = -\frac{1}{n-1}\Big\{ \sum_{t=2}^{T} u_t - (\hat{\rho}_n - \rho) \sum_{t=2}^{n} y_{t-1} \Big\}$$

且 $\eta(n) = O_P(n^{-1} a_n)$。

证明 由定义

$$\begin{aligned} u_k^0 - u_k &= \hat{u}_k - \frac{1}{n-1}\sum_{k=2}^{n} \hat{u}_k - u_k \\ &= [y_k - \hat{\rho}_n y_{k-1}] - \frac{1}{n}\sum_{k=2}^{n} [y_k - \hat{\rho}_n y_{k-1}] - u_k \\ &= \big[y_k - \rho y_{k-1} - (\hat{\rho}_n - \rho) y_{k-1}\big] - \frac{1}{n-1}\sum_{k=2}^{n} \big[y_k - \rho y_{k-1} + (\rho - \hat{\rho}_n)\big] - u_k \\ &= -(\hat{\rho}_n - \rho) y_{k-1} - \frac{1}{n-1}\Big\{ \sum_{k=2}^{n} u_k - (\hat{\rho}_n - \rho) \sum_{k=2}^{n} y_{k-1} \Big\} \\ &= \gamma_k + \eta(n), \end{aligned}$$

而由引理1可知，$\eta(n)=O_P(n^{-1}a_n)$，从而引理证毕。

引理7　在定理3的条件下，对任意x，均有

$$\lim_{n\to\infty}\frac{1}{n-b}\sum_{k=2}^{n-b+1}P(T_{b,k}^0\le x)\to H(x).$$

证明　与引理4的证明类似，仅需在备则假设下证明下面两式成立

$$\limsup_{n\to\infty}\max_{2\leqslant k\leqslant n-b+1}P\big\{\,|R_{b,k}^N|>\varepsilon\,\big\}=0,\quad \limsup_{n\to\infty}\max_{2\leqslant k\leqslant n-b+1}P\big\{\,|R_{b,k}^N|>\varepsilon\,\big\}=0.$$

以下仅证明第一项，第二项可类似证明。在引理4与引理6的相关记号下，有

$$R_{b,k}^N=T_{b,k}^{0N}-T_{b,k}^N=U-V+X-Y+Z,$$

其中

$$\begin{aligned}
U&=b^{-4}a_b^{-2}(a_{b,k}c_{b,k}-b_{b,k}^2)Y'E_{b,k}^0-b^{-4}a_b^{-2}(a_{b,k}c_{b,k}-b_{b,k}^2)Y'E_{b,k}\\
&=O(b^{-4})a_b^{-2}[Y'E_{b,k}^0-Y'E_{b,k}],
\end{aligned}$$

$$\begin{aligned}
&a_b^{-2}[Y'E_{b,k}^0-Y'E_{b,k}]\\
=\;&a_b^{-2}\Big(\sum_{i=k}^{k+b-1}y_{i-1}(k)u_i^0-\sum_{i=k}^{k+b-1}y_{i-1}(k)u_i\Big)\\
=\;&\Big(a_b^{-1}\sum_{i=k}^{k+b-1}u_i\Big)\Big[a_b^{-1}\sum_{i=k}^{k+b-1}\big(\eta(n)+\gamma_i\big)\Big]+\frac{1}{2}a_b^{-2}\Big[\sum_{i=k}^{k+b-1}\big(\eta(n)+\gamma_i\big)\Big]^2\\
&+a_b^{-2}\sum_{i=k}^{k+b-1}u_i\big(\eta(n)+\gamma_i\big)-\frac{1}{2}a_b^{-2}\sum_{i=k}^{k+b-1}\big(\eta(n)+\gamma_i\big)^2\\
=\;&\Big\{\Big(a_b^{-1}\sum_{i=k}^{k+b-1}u_i\Big)\Big(a_b^{-1}\sum_{i=k}^{k+b-1}\eta(n)\Big)+\frac{1}{2}a_b^{-2}\sum_{i=k}^{k+b-1}\eta(n)+a_b^{-2}\sum_{i=k}^{k+b-1}u_i\eta(n)-\frac{1}{2}a_b^{-2}\sum_{i=k}^{k+b-1}\eta^2(n)\Big\}\\
&+\Big\{-a_b^{-2}\sum_{i=k}^{k+b-1}\eta(n)\gamma_i-\frac{1}{2}a_b^{-2}\sum_{i=k}^{k+b-1}\gamma_i^2\Big\}\\
&+\Big\{a_b^{-1}\sum_{i=k}^{k+b-1}\eta(n)\gamma_i+\frac{1}{2}a_b^{-1}\sum_{i=k}^{k+b-1}\gamma_i^2-\Big(a_b^{-1}\sum_{i=k}^{k+b-1}u_i\Big)a_b^{-1}\sum_{i=k}^{k+b-1}\gamma_i+a_b^{-2}\sum_{i=k}^{k+b-1}u_i\gamma_i\Big\}\\
=\;&U_{b,k}^{N1}+U_{b,k}^{N2}+U_{b,k}^{N3}.
\end{aligned}$$

由于当$n\to\infty$时，$b/n\to\infty$，结合引理6可得

$$\limsup_{n\to\infty}\max_{1\leqslant k\leqslant n-b+1}P\big\{\,|U_{b,k}^{N1}|>\varepsilon\,\big\}=0,$$

而由文献[14]中引理6的证明可得

$$\limsup_{n\to\infty}\max_{1\leqslant k\leqslant n-b+1}P\big\{\,|U_{b,k}^{N2}|>\varepsilon\big\}=0,\quad \limsup_{n\to\infty}\max_{1\leqslant k\leqslant n-b+1}P\big\{\,|U_{b,k}^{N3}|>\varepsilon\,\big\}=0,$$

综合上述，则有 $U = o_P(1)$，类似可证 $V = o_P(1)$，$X = o_P(1)$，$Y = o_P(1)$，$Z = o_P(1)$，从而

$$\limsup_{n\to\infty} \max_{2\leqslant k\leqslant n-b+1} P\{\,|R_{b,k}^N| > \varepsilon\,\} = 0,$$

故引理证毕。

引理8 在定理3的条件下，对任意 x，均有 $\mathrm{Var}[\hat{G}_b(x)] \to 0$。

证明 证明与引理5相同，故略。

定理3的证明 为证 $\hat{G}_b(x) \xrightarrow{P} G(x)$，仅需证明 $E[\hat{G}_b(x) - G(x)] \to 0$, $\mathrm{Var}[\hat{G}_b(x) - G(x)] \to 0$。而由引理7、8可以得到 $E[\hat{G}_b(x) - G(x)] \to 0$, $\mathrm{Var}[\hat{G}_b(x) - G(x)] \to 0$，从而定理证毕。

4 模拟研究

本节用模拟研究来验证本文中subsampling方法的有效性。考虑如下数据生成过程(DGP)

$$y_t = \mu + \theta DU_t + \rho y_{t-1} + u_t, \quad DU_t = \begin{cases} 1, & t \le k_0, \\ 0, & t > k_0, \end{cases} \quad t = 1, \cdots, n, \tag{9}$$

$y_0 = 0$，其中参数 $\mu = 0.1$, $\theta = 0.2$，DU_t 如前所定义，u_t 为厚尾指数 $\kappa = 1.14$ 的独立随机变量序列。

表1为在原假设下，即 $\rho = 1$，样本量 $n = 1000$ 时，直接模拟重复5000次试验所得百分位数，表2为样本量 $n = 1000$ 时，子样本量为 $m = [\ln(n)^2]$，重复抽样5000次时所得百分位数。对比表1与表2，subsampling方法确定的百分位数波动较小，且计算运行时间较短。表3与表4分别为在备择假设下，$\rho = 0.7$，利用表1与表2中的分位数计算所得经验势函数值，重复试验次数为5000次。对比表3与表4，subsampling抽样所得临界值的势与直接模拟所得临界值相近，同时，当样本量 n 逐渐增大时，两种方法的经验势函数均接近于1。

表1: 当 $\kappa = 1.14$ 时，Monte Carlo模拟所得渐近分布(2)的百分位数

	1%	2.5%	5%	10%	90%	95%	97.5%	99%
λ=0.1	-49.6916	-33.0439	-25.8829	-20.6187	-4.8297	-3.7336	-2.9012	-1.8471
0.2	-55.8427	-36.2811	-26.9334	-21.0022	-5.2803	-4.0516	-3.0127	-1.5167
0.3	-57.3373	-36.4277	-26.9860	-21.1285	-4.8980	-3.5957	-2.4422	-0.8428
0.4	-53.4348	-36.7738	-27.8241	-20.9102	-4.3967	-3.1490	-2.1149	-0.6182
0.5	-54.7308	-35.7396	-26.9451	-20.8586	-3.9839	-2.9013	-1.7961	-0.4268
0.6	-52.9663	-38.3679	-28.2609	-21.2167	-4.4016	-3.1920	-1.9892	-0.6271
0.7	-64.5498	-37.2638	-28.0128	-21.5163	-5.0157	-3.8594	-2.6479	-1.0842
0.8	-60.3578	-36.6004	-27.8867	-21.4129	-5.3179	-4.0841	-3.0800	-1.8164
0.9	-52.3821	-34.7184	-26.6580	-20.3535	-4.8968	-3.7787	-2.8314	-1.4228

表 2:　当 $\kappa = 1.14$ 时，subsampling 方法模拟所得渐近分布 (2) 的百分位数

	1%	2.5%	5%	10%	90%	95%	97.5%	99%
λ=0.1	-45.1578	-25.7280	-21.4191	-16.7911	-5.0268	-3.6300	-2.3096	-1.7618
0.2	-42.2333	-29.1284	-23.0033	-18.2148	-5.4403	-4.3610	-3.4893	-1.3655
0.3	-43.3304	-30.2418	-23.0136	-18.6540	-5.2126	-4.2637	-3.6270	-2.1643
0.4	-44.3758	-30.7756	-24.0014	-18.8300	-4.6412	-3.8584	-2.6047	-1.6014
0.5	-45.1691	-36.9219	-30.4722	-22.9231	-4.7443	-4.0289	-2.9999	-1.7521
0.6	-45.1691	-36.9219	-30.5058	-22.9231	-4.7055	-3.9355	-2.7280	-1.0672
0.7	-45.1691	-36.9219	-30.4722	-22.5792	-4.8835	-3.8678	-2.7280	-1.0672
0.8	-45.1691	-32.2741	-28.8533	-21.5030	-4.8835	-4.0178	-2.6694	-1.5482
0.9	-45.1691	-31.6323	-26.5962	-20.7075	-4.8404	-3.8822	-2.5633	-0.9671

表 3:　当 $\kappa = 1.14$ 时，不同样本容量时 Monte Carlo 模拟所得经验势函数

	$n = 100$			$n = 200$			$n = 300$		
λ	1%	2.5%	5%	1%	2.5%	5%	1%	2.5%	5%
0.1	0.0536	0.6524	0.9516	0.9764	0.9998	1.0000	0.9996	0.9998	0.9998
0.2	0.0126	0.4602	0.9430	0.9162	0.9996	0.9998	0.9994	0.9998	0.9998
0.3	0.0092	0.4474	0.9394	0.8970	0.9994	0.9996	0.9992	1.0000	1.0000
0.4	0.0186	0.4048	0.9272	0.9526	0.9990	1.0000	0.9996	1.0000	1.0000
0.5	0.0156	0.4782	0.9364	0.9364	0.9994	0.9998	0.9996	1.0000	1.0000
0.6	0.0224	0.3060	0.9182	0.9564	0.9980	0.9994	0.9996	1.0000	1.0000
0.7	0.0044	0.3866	0.9204	0.5674	0.9998	0.9998	0.9970	0.9998	0.9998
0.8	0.0058	0.4246	0.9180	0.8024	0.9998	1.0000	0.9986	1.0000	1.0000
0.9	0.0392	0.5296	0.9380	0.9598	0.9990	0.9998	0.9998	1.0000	1.0000

表 4:　当 $\kappa = 1.14$ 时，不同样本容量时 subsampling 方法所得经验势函数

	$n = 100$			$n = 200$			$n = 300$		
λ	1%	2.5%	5%	1%	2.5%	5%	1%	2.5%	5%
0.1	0.1882	0.9544	0.9902	0.9968	1.0000	1.0000	1.0000	1.0000	1.0000
0.2	0.1542	0.8884	0.9864	0.9970	1.0000	1.0000	1.0000	1.0000	1.0000
0.3	0.1280	0.8618	0.9868	0.9936	0.9994	0.9996	1.0000	1.0000	1.0000
0.4	0.0932	0.8300	0.9792	0.9938	0.9994	0.9996	1.0000	1.0000	1.0000
0.5	0.0838	0.3876	0.8464	0.9922	0.9990	0.9998	0.9998	0.9998	0.9998
0.6	0.0818	0.3974	0.8388	0.9920	0.9992	0.9996	0.9998	0.9998	0.9998
0.7	0.0886	0.4174	0.8500	0.9924	0.9996	0.9998	1.0000	1.0000	1.0000
0.8	0.0896	0.7396	0.9000	0.9914	0.9994	0.9996	1.0000	1.0000	1.0000
0.9	0.1050	0.7596	0.9430	0.9914	0.9998	0.9998	1.0000	1.0000	1.0000

参考文献：

[1] Perron P. The great crash, the oil price shock and the unit root hypothesis[J]. Econometrica, 1989, 57: 1361-1401

[2] Perron P. Testing for a unit root in time series with a changing mean[J]. Journal of Business and Economic Statistics, 1990, 8: 153-162

[3] Banerjee A, *et al.* Recursive and sequential tests of the unit root and trend break hypothesis: theory and international evidence[J]. Journal of Business and Economic Statistics, 1992, 10: 271-287

[4] Christiano L. Searching for breaks in GNP[J]. Journal of Business and Economic Statistics, 1992, 10: 237-250

[5] Zivot E, Andrews D W K. Further evidence on the great crash, the oil price shocks, and unit root hypothesis[J]. Journal of Business and Economic Statistics, 1992, 10: 251-270

[6] Kim T, Leybourne S, Newbold P. Unit root tests with a break in innovation variance[J]. Jouranl of Econometrics, 2002, 109: 365-387

[7] Nunes L. LM-type tests for a unit root allowing for a break in trend[C]// Econometric Society 2004 Austrlasian Meetings, Econometric Society, 2004, No.190

[8] Guillaume D M, *et al.* From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets[J]. Finance and Stochastics, 1997, 1: 95-129

[9] Mittnik S, Rachev S. Stable Paretian Models in Finance[M]. New York: Wiley, 2000

[10] Davis R A, Resnick S. Limit theory for moving averages of random variables with regularly varying tail probabilities[J]. The Annals of Probability, 1985, 13(1): 179-195

[11] Davis R A, Resnick S. Limit theory for the sample covariance and correlation functions of moving averages[J]. The Annals of Statistics, 1986, 14(2): 533-558

[12] Chan N H. On the first order autoregressive process with infinite variance[J]. Econometric Theory, 1989, 5: 354-362

[13] Phillips P C B. Time series regression with a unit root and infinite-variance errors[J]. Econometric Theory, 1990, 6: 44-62

[14] Agnieszka J, Kokoszka P. Subsampling unit root tests for heavy tailed observations[J]. Methodology and Computing in Applied Probability, 2004, 6(1): 73-97

[15] 陆懋祖. 高等时间序列经济计量学[M]. 上海：上海人民出版社，1999
Lu M Z. Advanced Time Series Econometrics[M]. Shanghai: Shanghai People's Publishing House, 1999

Subsampling Procedures for the Heavy-tailed Unit Root Test with Structural Change

QIN Rui-bing[1], TIAN Zheng[1,2]

(1- Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an 710072;
2- State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications,
Chinese Academy of Science, Beijing 100101)

Abstract: This paper deals with the unit root test for the series with the infinite-variance innovations and the change in trend. The DF-type test is extended and the asymptotic distribution is obtained. In order to get the percentiles of the DF-type test, subsampling procedures are designed, and the consistency of the subsampling procedure is established. Finally, Monte Carlo simulations demonstrate the validity of the proposed test and the subsampling procedures in this paper.

Keywords: processes with infinite variances; change point; unit root test; subsampling procedures

Received: 07 July 2009. **Accepted:** 13 Jan 2010.

Foundation item: The National Natural Science Foundation of China (60972150; 10926197; 60375003); the Science and Technology Innovation Foundation of Northwestern Polytechnical University (2007KJ01033).